DOCUMENT RESUME

ED 343 920 TM 018 061

AUTHOR Cariaga-Lo, Liza; And Others
TITLE Self-Assessment in Medical Students: Consistency on Short-Essay Subject Matter Examinations.
PUB DATE 92
NOTE 13p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Comparative Testing; *Essay Tests; Feedback; Higher Education; Longitudinal Studies; Medical Evaluation; *Medical Students; *Predictive Measurement; *Science Tests; Self Concept; *Self Evaluation (Individuals; Student Evaluation
IDENTIFIERS Accuracy; Rater Reliability

ABSTRACT

Results from an ongoing series of studies designed to describe changes in students' self-assessments across time are presented. In these studies, medical students predicted their performance on essay examinations in their first 2 years of medical school. Self-assessment is a valuable skill for clinical practice, and the importance of knowing what one knows is a reason for giving students feedback on their self-assessment. Data from two classes in a problem-based curriculum (one class of 22 students; other class size not give) for four 6-hour essay examinations taken in students' first two years show different patterns in accuracy of self-estimation. The class of 1993 improved from the first to the third examination, with a decline in accuracy for the fourth examination. Data from the class of 1994 show an up and down trend. These differences may result from intrinsic factors affecting students' self-assessments over time. Qualitative analyses will be needed to determine the reasons for the changing accuracy of self-assessment. Included are six tables of study data and a six-item list of references. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LIZA CARIAGA-LO
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED343920

Self-Assessment in Medical Students: Consistensy on Short-Essay Subject Matter Examinations

Liza Cariaga-Lo, Ann W. Frye, Boyd F. Richards, Elizabeth W. Bradley, and James R. Philp.
Bowman Gray School of Medicine of Wake Forest University

> "Self-assessment refers to the involvement of the learners in making judgments about their own learning, particularly about their achievements and the outcomes of their learning. Self-assessment is formative in that it contributes to the learning process and assists learners to direct their energies to areas for improvement, and it may also be summative, either in the sense of learners deciding that they have learned as much as they wished to in a given area, or, in formal institutional settings, it may contribute to the grades awarded to students." (Boud & Falchikov, 1989, p.529).

Self-assessment is an important component of learning in that it acts as a regulatory mechanism by which students can gauge the depth of their knowledge in a subject. In the field of medicine, accurate self-assessment is a particularly important dimension of clinical practice. Overestimation can lead to premature closure, misdiagnosis or inadequate performance of procedures. Underestimation can lead to excessive uncertainty, overuse of diagnostic tests, and unnecessary referrals to other health professionals. Thus, the importance of accurately knowing what one knows provides a compelling reason to evaluate and give feedback to medical students about the opportunity to take a proactive role in their learning process. How can we begin to accurately evaluate students' abilities to make self-assessments about their learning?

TM018061

BEST COPY AVAILABLE

OBJECTIVES:

The paper discussed today will report results from an on-going series of studies in which students predict their performance on essay exams during their first two years of medical school. The objective of these studies is to describe the changes in students' self-assessments across time, give the same kind of self-assessment task on the same type of examinations covering the same disciplines. By controlling for important external factors (i.e., type of examination, type of performance standard, and complexity of performance domain), we hope to demonstrate the presence of an intrinsic tendency which influences self-assessment accuracy. This approach is consistent with recent longitudinal self-assessment studies (Arnold, Willoughby and Calkins, 1985). To summarize, the objective of this paper will be:

1. Specifically discuss the manner in which we have evaluated students' self-assessments over time.
2. Interpret these findings to argue for the presence of an intrinsic tendency in self-assessments among medical students in a problem-based curriculum.
3. Discuss how our methodology could be better informed by more qualitative methods of evaluating self-assessment.

LITERATURE BACKGROUND:

Medical educators have experimented during the past 20 years with a variety of techniques to evaluate students' self-assessment skills. (see Falchikov and Boud (1989) for a review of

this literature. The most frequent method of evaluation generally involves comparison of students' self-assessments with other measures of their actual performance, usually faculty or peer ratings. Self-assessment techniques consistent with this approach have ranged from broad estimates of overall performance on an exam or clerkship (Kolm and Verhulst, 1987; Arnold, et al.,1985) to very precise estimates about the probability of alternatives on multiple choice tests being correct (Voytovich and Rippey, 1982). Self-assessment studies published since 1970 have focused on describing the accuracy of a group of students predicting their performance on a single evaluation measure. While these studies have produced mixed results, most found that students, on the whole, tended to rate themselves lower than did faculty or peers (Morton and Macbeth, 1977; Stuart, Goldstein and Snope, 1980; Calhoun, Ten Haken and Woolliscroft, 1990; Peters and Schimpfhauser, 1990). Additional studies have shown that individual self-assessments vary dramatically within a group and often change between performance situations and across time (Arnold, et.al., 1985; Falchikov and Boud, 1989). What can account for these differences?

Research findings suggest several possible interacting factors ranging from a student's intrinsic self-assessment tendency to the external characteristics of a given performance situation. Keefer (1971) found that students who made accurate self-assessments, compared to students who made inaccurate ones, tended to be upperclassmen, to have higher academic achievement

scores, and to be more successful in achieving their goals. In addition, Falchikov and Boud (1989) found that students' familiarity with the performance domain being assessed, the domain's complexity and the clarity and specificity of the performance standard all appeared to influence the accuracy of students' self-assessments.

METHODS:

All medical students enrolled in the problem-based medical school curriculum participate in these studies. Thus far, we have data from the classes of 1993 and 1994. The students in the problem-based curriculum had self-selected this curriculum upon admission to medical school. However, they resembled their classmates in the traditional lecture-based curriculum with respect to age, race and preadmissions academic measures.

The students complete four six-hour essay exams at intervals during their first two years of medical school. The exams each contain an average of 40 basic science questions related to several clinical scenarios. An interdisciplinary committee of basic scientists and clinicians write the questions for each exam. A faculty expert then scores all answers to a single question using a 0-4 scale with intervals of 0.25. A self-assessment sheet is attached to each student's test booklet at the time of the exam. This contains a 0-4 printed scale for each question on which students were directed to estimate their performance.

Feedback to the students consists of their scores on each

question answered, a protocol answer for each question, and a profile of their composite scores by discipline.

For analysis of the data collected to date (two complete classes), we converted the experts' scores and students' self-assessments to percentages. We then calculated the difference between students' prediction and the experts' scores for each question (e.g., a score of 3.0 out of a possible 4.0 was converted to 75%). The degree to which students had over- and under-estimated their performance across all questions was determined by separately averaging all positive and negative differences between students' and experts' assessments on each exam.

We then looked at individual means and standard deviations by phase for each class. We used one-way ANOVA to examine any significant differences in averages (of scores, accuracy, overestimation, and underestimation) by phase for each class.

RESULTS:

The data analyzed for the Class of 1993 indicate that accuracy in self-assessment improves from Exam 1 to Exam 3 (with less overestimation) and drops on Exam 4 (with more underestimation).(See Table 2). In contrast, data from the Class of 1994 had no such tendencies. There seemed to be an up and down trend from exam to exam in the Class of 1994. (See Table 3). Although the examination given from year to year and the environmental factors from year to year were similar for both

classes, there were indeed differences in self-assessment trends between the two years. For the 1993 class, total scores given by faculty from exam to exam were significantly different. For the 1994 class, total scores given by faculty remained stable from exam to exam. Both classes, however, indicated that accurate estimations differed from exam to exam. How can we explain these differences?

DISCUSSION:

These differences in accurate estimations within each class may be a result of intrinsic factors affecting medical students' self-assessments over time. The differences in over- and under-estimation trends between classes could be a function of different intrinsic factors operating for each class group. Perhaps there are differences in achievement motivation, student maturity level, student self-confidence, and other such intrinsic factors. However, quantitative analyses alone cannot give justice to the complexity of students' self-assessment behaviors. Qualitative analyses must be carried out in studies seeking to elucidate the nature of self-assessment behaviors in students. For instance, we are currently in the midst of starting a qualitative study, using these same students. We argue that interview questions and administered self-assessment tasks can more refinedly get at the process of self-assessment in students. Such triangulation of methods can give us greater insight into what factors (external or intrinsic) specifically affect student self-assessment over time. In doing so, we can then give better

feedback to students about their ability to judge their learning performance.

CONCLUSIONS:

The quantitative findings from two medical school classes have indicated that there are probably intrinsic factors affecting self-assessment in basic science content exams. However, qualitative studies, in conjunction with quantitative studies, can get us a step closer to forming a comprehensive model of self-assessment behaviors. Thus far, no comprehensive model of self-assessment behaviors has been established in the literature. The quantitative findings shown here today is a necessary first step towards this process.

The ability of individuals to master new areas of knowledge can be affected by their ability to judge their level of performance and progress in these areas of knowledge. Thus, self-assessment is an important skill to be developed by students, particularly with the explosion of new and increasingly more complex knowledge base in all domains.

Further research in this area needs to investigate the generalizability of these findings using similar methodologies with other student groups and in other performance areas. In doing so, we can then be more cognizant of how self-assessment may affect the quality of learning for an individual.

"One of the characteristics of effective learners is that they have a realistic sense of their own strengths and weaknesses and that they can use knowledge of their own achievements to direct their studying into productive directions..."Good" students have always been effective self-assessors, but it is becoming increasingly recognised that in order to develop this skill more widely among students, explicit attempts need to be openly practised. Self-assessment has also been associated with moves towards developing greater student autonomy in learning, and particularly self-directed learning..."(Falchikov & Boud, 1989, p.530).

Table 1:

	BACKGROUND INFORMATION		
	UGrad School	UGrad Major	Total MCAT
Student A	UNC	NonSC	49
Student B	Yale	Biology	43

Table 2:

	TOTAL POINTS RECEIVED ON THE ESSAY EXAMS		
	Exam 1	Exam 2	Exam 3
Student A	63%	52%	49%
Student B	64%	55%	47%
Class (n=22)	71% sd=25%	60% sd=25%	59% sd=25%

Table 3:

	POINTS NOT PREDICTED BUT RECEIVED (UNDER-ESTIMATION)*		
	Exam 1	Exam 2	Exam 3
Student A	1%	5%	10%
Student B	.5%	7%	6%
Class	7% sd=13%	7% sd=13%	8% sd=13%

Table 4:

	POINTS PREDICTED BUT NOT RECEIVED (OVER-ESTIMATION)*		
	Exam 1	Exam 2	Exam 3
Student A	21%	13%	6%
Student B	32%	23%	13%
Class	15% sd=20%	14% sd=17%	11% sd=16%

* Key to interpreting under- and over-estimation scores:

Total Points Possible on an Exam			
Predicted but NOT received (over-estimation)	Predicted and received (accurate)	NOT predicted but received (under)	NOT predicted and NOT received (accurate)

<----------Points Predicted---------->

<----------Points Received---------->

Table 2
CLASS OF 1993
DEPENDENT MEASURES BY EXAM

	EXAM 1	EXAM 2	EXAM 3	EXAM 4
Means and SDs (% of total points on each exam):				
Total Score*	70.80	59.90	59.44	61.76
	7.37	11.69	12.01	9.62
Accurate Estimation*	78.13	80.11	81.61	79.97
	4.46	4.74	4.10	2.38
Under Estimation*	7.27	6.73	7.46	11.53
	6.45	4.89	3.52	4.46
Over Estimation**	14.60	13.17	10.93	8.50
	7.15	7.21	4.76	4.20

* Significant One-way ANOVA ($p<.05$)
** Significant One-way ANOVA ($p<.05$)

Table 3
CLASS OF 1994
DEPENDENT MEASURES BY EXAM

	EXAM 1	EXAM 2	EXAM 3	EXAM 4
Means and SDs (% of total points on each exam):				
Total Score	60.60 8.23	59.36 7.01	59.84 15.82	58.04 9.37
Accurate Estimation*	76.66 3.01	80.23 5.07	76.11 5.16	79.99 5.16
Under Estimation	15.08 7.19	12.76 8.12	17.17 10.12	11.42 7.45
Over Estimation	8.26 5.57	7.01 4.95	6.72 7.89	8.59 5.53

* Significant One-way ANOVA ($p<.05$)

References

Arnold, L., Willoughby, T.L., & Calkins, E.V. (1985) Self-evaluation in undergraduate medical education: a longitudinal perspective. Journal of Medical Education 60, 21-28.

Calhoun, J.G., Ten Haken, J.D. & Woolliscroft, J.O. (1990) Medical students' development of self- and peer-assessment skills: a longitudinal study. Teaching and Learning in Medicine, 2, 25-29.

Falchikov, N. & Boud, D. (1989) Student self-assessment in higher education: a meta-analysis. Review of Educational Research, 59, 395-430.

Keefer, K.E. (1971) Characteristics of students who make accurate and inaccurate self-predictions of college achievement. The Journal of Educational Research, 64, 401-404.

Kolm, P. & Verhulst, S.J. (1987) Comparing self- and supervisor evaluations: a different view. Evaluation & the Health Professions, 10, 80-89.

Morton, J.B. & Macbeth, W.A. A. G. (1977) Correlations between staff, peer and self assessments of fourth-year students in surgery. Medical Education, 11, 167-170.

BEST COPY AVAILABLE